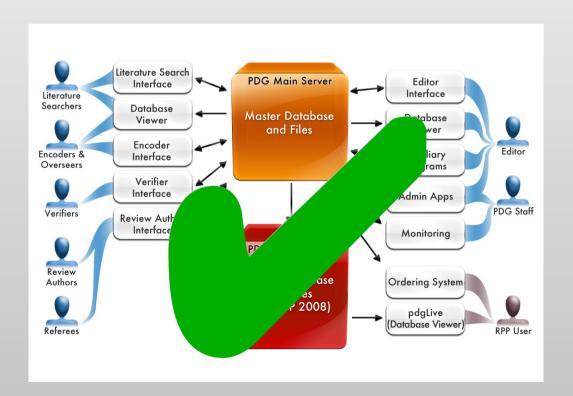




Summary of PDG Computing Upgrade

Juerg Beringer

Lawrence Berkeley National Laboratory

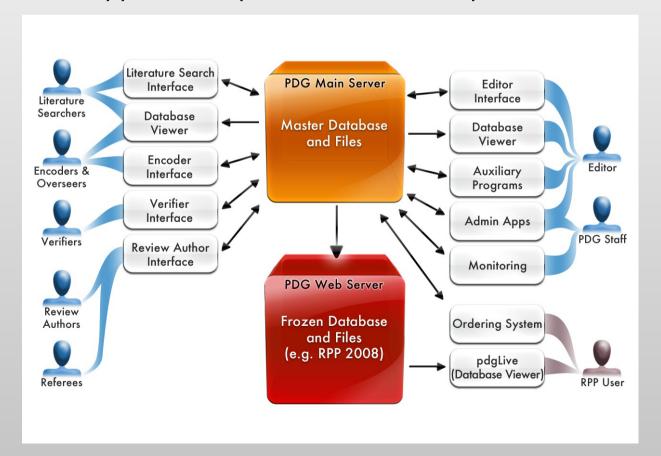




Goals and Scope



- Developed and proposed plan for comprehensive upgrade of PDG computing infrastructure in 2007
 - Primary goal: ensure PDG can continue to function well
 - Support all areas of PDG work, provide platform for future extensions
 - Set of web applications plus command-line production tools





Brief History



- Project funded following DOE review of PDG in 2008
 - 2 FTE for 3 years for development
 - 0.5 FTE (ongoing) for maintenance (starting in FY11)
- Backbone of new system deployed into production in 2010
- DOE review of computing upgrade project on 9/17/2010
 - "... upgrade is proceeding on schedule and within cost ..."
- Very positive feedback at PDG meetings in Fall 2010
- Presentation/publication at CHEP'2010 (Taiwan)
- Had hoped to complete project early (in August 2011)
 - Unexpected reduction in developer availability from May 2011 onwards meant we had to spread duration over full 3 years
 - Limited use of PdgWorkspace for 2012 edition
- Project completed on schedule and budget in December 2011

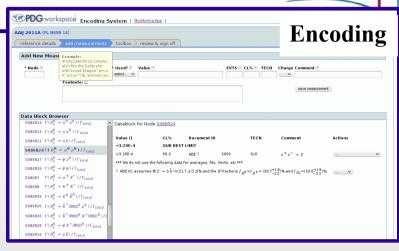


Results

- All primary / high-priority web applications plus many secondary ones completed
 - Very good progress on further features
- Achieved much more regarding database modernization than thought possible
- Modern, extendable, maintainable and well-documented system
- Will only discuss subset here









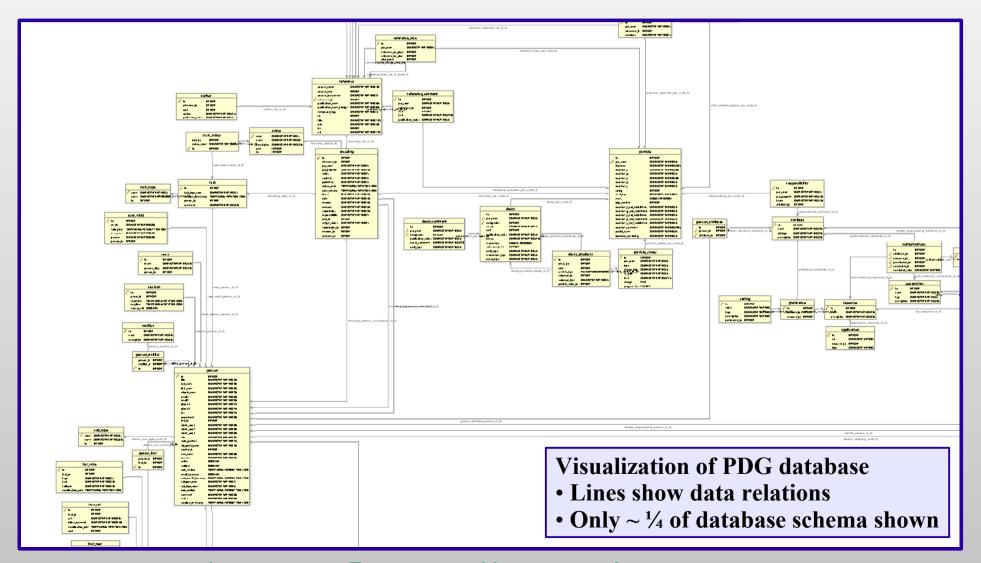




Modern PDG Database



- Large fraction of effort went into modernizing PDG database
 - One of the most valuable results of the project in the long term





PDG Workspace



- Framework used by all web applications for PDG collaborators
 - This is the starting point for working with the new system
- When people log in, they'll see a set of applications tailored to the work they do for PDG

PDG workspace						
Please log in.						
user name or email address: password:	login Forgot user name or password					

PDG wo

Ordering Admin App |

Monitoring App | Review App | Encoding System

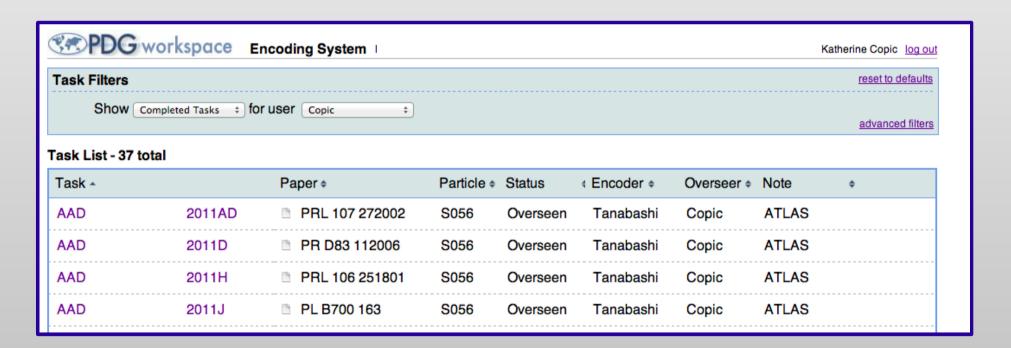
Juerg Beringer log out



Encoding System



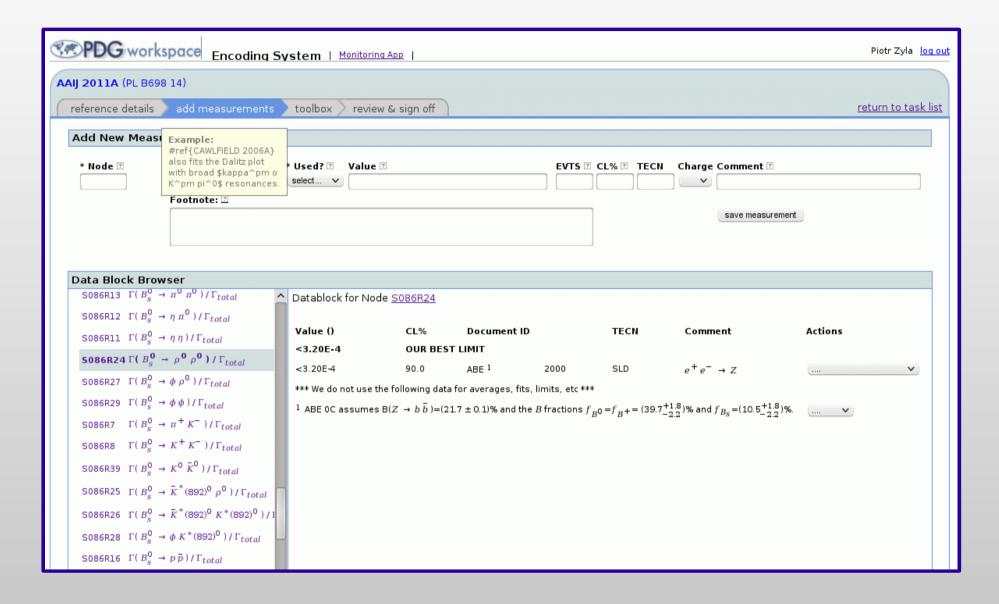
- Tool for updating the Listings
- Successfully used by Kathy Copic for all her 2012 encodings
 - Even being new to PDG, she found system very easy to use
- Gradually invite encoders/overseers to start using new system
 - Kathy Copic is contact person in case of problems





Data Entry Page







Review Authoring System



Tool for authoring and refereeing review articles

- Supports 3 ways for authors to write reviews
 - Traditional: check out tar file of sources, run TeX locally, return revised source files to overseer / editor, who updates them in system
 - Using subversion (SVN): check out source files via SVN, commit changes when ready, run TeX either locally or get PDF file from review authoring system
 - Online: edit source files in your browser, no local TeX needed

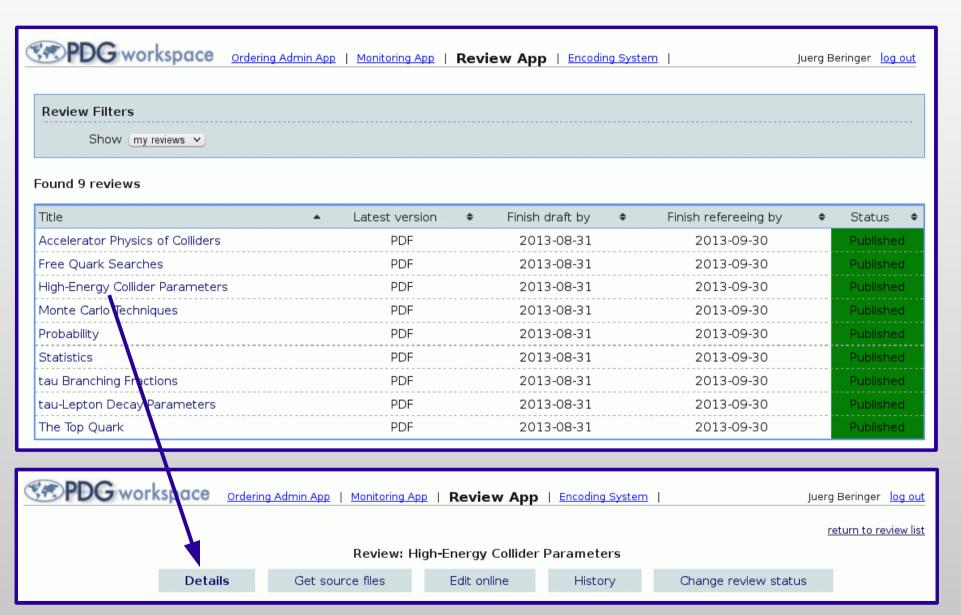
Other features include

- Keeps track of review responsibilities, status and deadlines
- History of changes and "diff" between versions in various forms
- Links between reviews and related items (reviews, data blocks) in order to improve consistency across RPP
- Integrated into overall progress monitoring
- Will be used starting with 2013 review updates



Review Task List



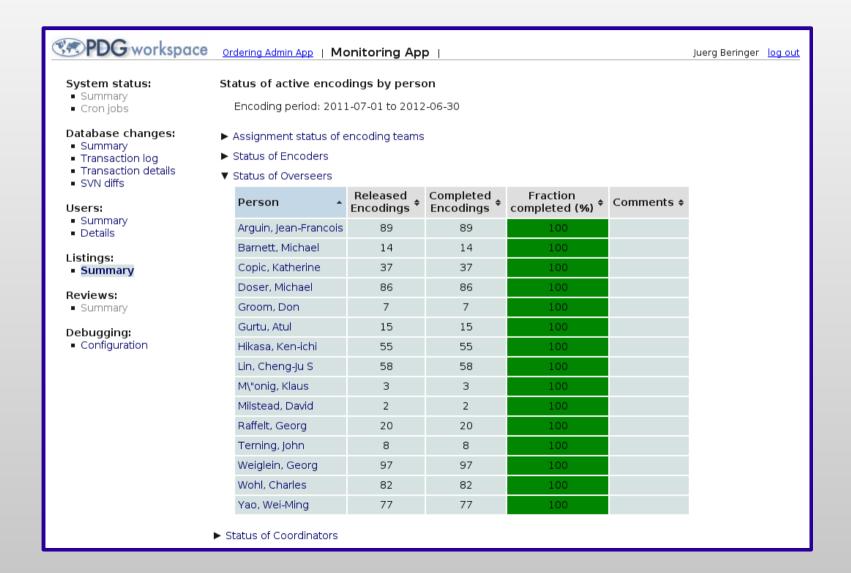




Monitoring: from Overall Progress



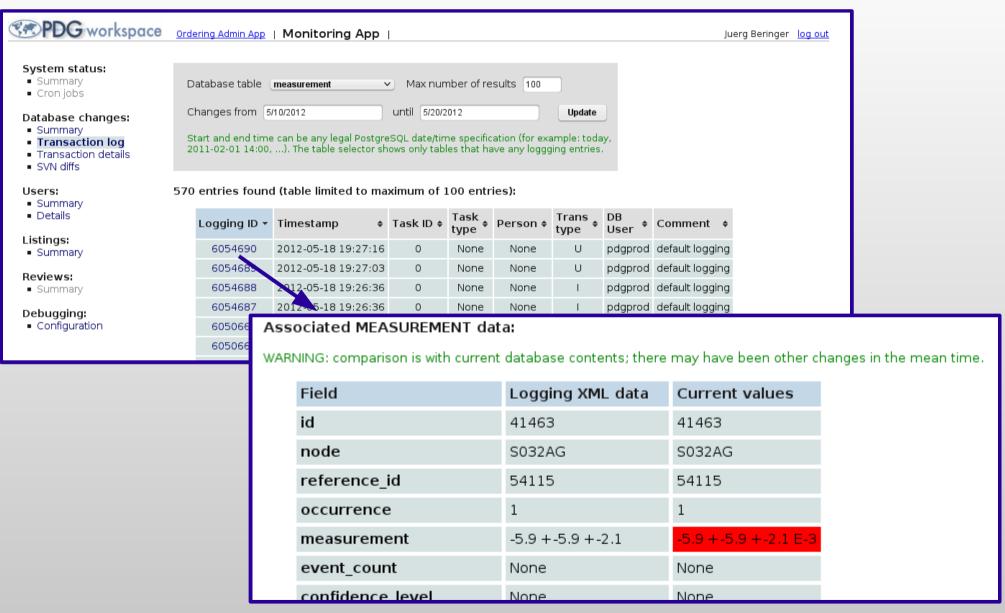
Used for 2012 edition and found to be very helpful





... to Detailed Database Changes



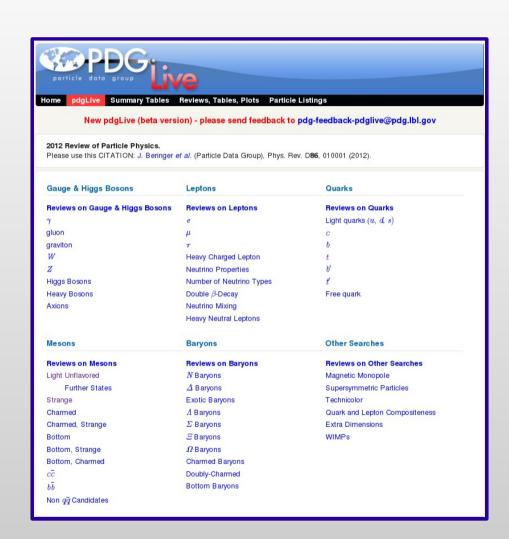




pdgLive



- Same look-and-feel as original pdgLive
- Many improvements and new features
- Deployed as "beta version" for RPP 2012, in parallel with original pdgLive
 - Wanted to be cautious, since maintenance person went on maternity leave two weeks before deployment
 - Will become primary version of pdgLive after fixing a few known issues
 - Already receives 2 to 3 times as much traffic as original pdgLive





Many Improvements



INSPIRE search

Print-like-quality display of TeX on the web (using MathJax)

Old pdgLive

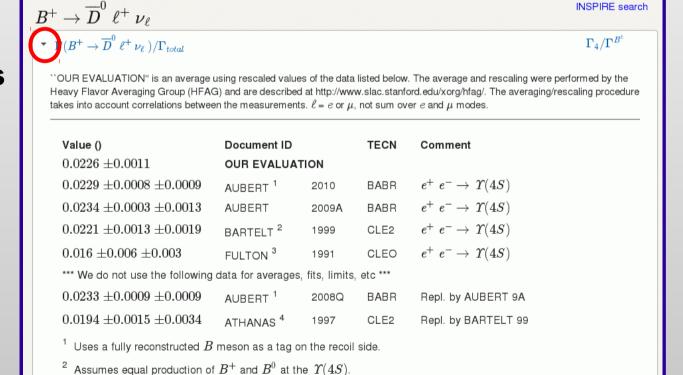
$$\Gamma_{6} \qquad \overline{D}^{*}(2007)^{0} I^{+} v_{I}$$

$$\Gamma_{7} \qquad \overline{D}^{*}(2007)^{0} \tau^{+} v_{I}$$

New pdgLive (beta)

$$egin{aligned} \Gamma_6 & B^+
ightarrow \overline{D}^* (2007)^0 \; \ell^+ \;
u_\ell \ & \Gamma_7 & B^+
ightarrow \overline{D}^* (2007)^0 \; au^+ \;
u_ au \end{aligned}$$

New presentation of branching ratios using folding sections





Many Improvements



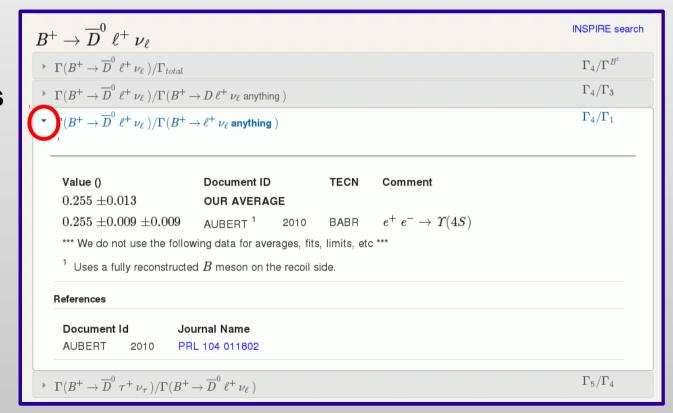
Print-like-quality display of TeX on the web (using MathJax)

Old pdgLive

New pdgLive (beta)

$$egin{aligned} oldsymbol{\Gamma}_6 & B^+
ightarrow \overline{D}^* (2007)^0 \; \ell^+ \;
u_\ell \ oldsymbol{\Gamma}_7 & B^+
ightarrow \overline{D}^* (2007)^0 \; au^+ \;
u_ au \end{aligned}$$

 New presentation of branching ratios using folding sections





Cross-Linking w/PDG Identifiers



Implemented with INSPIRE

Interest from Elsevier to do something similar in ScienceDirect





Searches w/PDG Identifiers





Home

pdgLive

Summary Tables

Reviews, Tables, Plots

Particle Listings

New pdgLive (beta version) - please send feedback to pdg-feedback-pdglive@pdg.lbl.gov

pdgLive Home >

2012 Review of Particle Physics.

Please use this CITATION: J. Beringer et al. (Particle Data Group), Phys. Rev. D86, 010001 (2012).

$$B^+
ightarrow \overline{D}^0 \ \ell^+ \
u_\ell \ oxedown_{\Gamma(B^+
ightarrow \overline{D}^0 \ \ell^+
u_\ell)/\Gamma_{total}}$$



INSPIRE search

$$\Gamma(B^+ \to \overline{D}^0 \ \ell^+ \ \nu_\ell \)/\Gamma_{total}$$

 Γ_4/Γ^{B^4}

"OUR EVALUATION" is an average using rescaled values of the data listed below. The average and rescaling were performed by the Heavy Flavor Averaging Group (HFAG) and are described at http://www.slac.stanford.edu/xorg/hfag/. The averaging/rescaling procedure takes into account correlations between the measurements. $\ell = e$ or μ , not sum over e and μ modes.

Value ()	Document ID		TECN	Comment
$0.0226\ {\pm}0.0011$	OUR EVALUA			
$0.0229\ {\pm}0.0008\ {\pm}0.0009$	AUBERT 1	2010	BABR	$e^+~e^- ightarrow \varUpsilon(4S)$
$0.0234 \pm 0.0003 \pm 0.0013$	AUBERT	2009A	BABR	$e^+~e^- ightarrow arUpsilon(4S)$



Ordering System



- Online since May 2011
- Will replace ordering via CERN library a few weeks after CERN ships 2012 booklets and diaries
 - Full statistics & contact info in one place (so far: split LBNL & CERN)
 - CERN will still ship orders for Europe, Africa, ...
- Users can subscribe to PDG mailing list(s)
- For PDG collaborators, same login as for PdgWorkspace





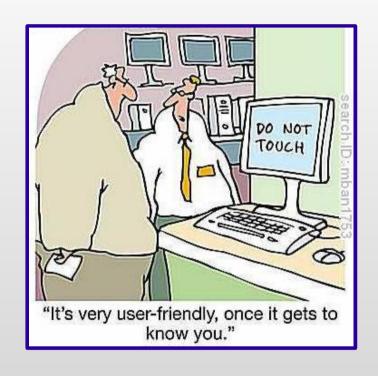
Start Broadly Using New System



- Want to gradually get collaborators started with new system
 - Allows to provide very good support and fix any problems quickly, if needed
- Start inviting encoders, overseers and literature searchers in December
 - Use new system for 2013 encodings
- Start inviting review authors in Spring/Summer 2013
 - In time for next review update



- Documentation (TWiki): http://pdgprod.lbl.gov/twiki/bin/view/Pdg/ (plus very detailed internal technical documentation)
- Each application has a responsible contact person see Twiki





New Possibilities



- New PDG computing system provides solid basis for implementing new features
 - Realizing substantial new features still requires a development effort
- Many ideas / requests for new features from PDG
 - Smartphone app
 - New features for pdgLive
 - Data entry by collaborations?
 - Download of PDG data in machine-readable format
 - For individual data blocks (XML, via pdgLive)
 - All PDG data (in a SQLite database)
 - Extending PDG computing platform to support averaging groups?
 - Programmatic access to PDG database outside of PDG collaboration?







Outcome of Discussions at Collaboration Meeting



- Most (but not all) collaborators opposed to allowing large collaborations (or authors in general) to suggest data entries
 - Even though PDG would validate/correct all entries published in RPP
 - Not much enthusiasm for a version of pdgLive that would, if the user chooses so, show contributed entries in addition to official PDG ones
 - Maybe let large collaborations provide references online
 - Reasons: increased amount of work (particularly in case of controversial issues), and the notion that this might undermine PDG's authority
- Downloadable data as a priority (anyway needed for PDG app)
- Cross-linking with HepData
- Plans and ideas from publishers reinforce our emphasis on developing both a PDG app and going beyond PDF format on the web (pdgLive, "article of the future", eBooks?, ...)
- Wish for author information in XML format and using ORCID



Conclusions



- PDG Computing Upgrade project successfully completed on schedule and within budget in December 2011
 - Backbone of new system used in production since summer 2010
 - Encoding interface and monitoring used for producing RPP 2012
- As a result of this project, PDG now has a solid computing platform on which we can build
 - Modern, extendable, maintainable and well-documented
- Gradually inviting collaborators to start using new system
 - Will allow us to provide very good support and fix problems quickly















Backup Slides



Intrinsically Challenging



- Convert ancient legacy software into modern system without disrupting ongoing PDG work
 - Gradually transform O(100) disconnected data tables into modern relational database with integrity constraints and change tracking
- Complex distributed workflow with hundreds of users
- Correctness of data is paramount
 - New system produces "identical" manuscript of 1400+ page book
- Display of math on the web
- PDG schedule leaves only few deployment opportunities
 - Gradual deployment, starting in August 2010



Our Approach



Focus and prioritize

- Identified technically challenging areas: database, math display
- Focus on cornerstone of new system: modern PDG database
- Complete primary and difficult applications / middleware first
- Secondary applications lower priority than publication-critical ones

Avoid scope creep and keep to original goals

 Sometimes hard, since new system offers so many exciting opportunities for extension beyond the project's scope

Strong involvement from PDG group

- Orin Dahl, Piotr Zyla, JB
- Many others helped with testing



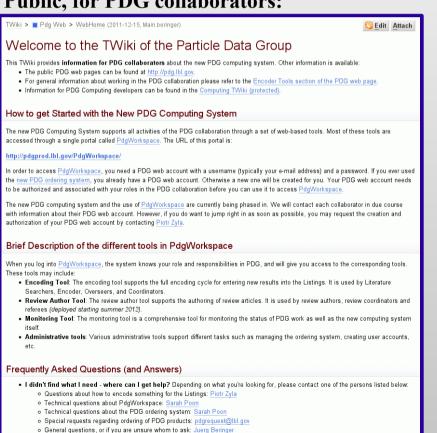
Documentation



Excellent documentation emphasized from beginning

- Manuals
- TWiki
- Inline code documentation, man pages, help

Public, for PDG collaborators:



Protected, for developers:





Knowledge Transfer



- Hands-on involvement of PDG personnel ensures excellent knowledge of new system from start
- Schema migration tutorial/exercise on 12/21/2011
 - Using a copy of the new system locally installed on our laptops, successfully modified the database schema, updated all necessary software components, and recompiled everything
- Sarah Poon retained as maintenance person
 - On maternity leave since 6/13/2012, back since September
- Other developers still accessible at LBNL if needed



Project Team



- Juerg Beringer (PDG physicist)
 - Project leader, requirements, system architecture
- Chuck McParland (computer scientist)
 - Java API
- Sarah Poon (computer systems engineer)
 - Web design, user interfaces, JavaScript

Each member of the team has many years of software development experience

- David Robertson (computer systems engineer)
 - Database, Python API, scripts
- Orin Dahl (PDG physicist, retired)
 - Legacy Fortran programs
- Piotr Zyla (PDG editor)
- Contributions from Jacob Andreas, Cecilia Aragon, Keith Beattie,
 Igor Gaponenko, Keith Jackson, Kirill Lugovsky, Slava Lugovsky



Automated Testing



```
from PDG Nightly Builds <pdg-build@george.lbl.gov>☆

subject PDG nightly build report

to PDG Build list <pdg-build@george.lbl.gov>☆

pdg-job-runner.sh: Wed May 30 04:01:01 PDT 2012 starting...

success pdgbuild-recreate.sh in 112 secs
success python.sh in 4 secs
success JavaApi in 39 secs
success JavaDao in 24 secs
success JavaDao in 24 secs
success AuthApi in 25 secs
success JavaModels in 23 secs
* failure PdgWorkspace (1) in 32 secs

* failure PdgWorkspace (1) in 32 secs

See <a href="http://pdg6.bl.gov/cgi-bin/pdg-monitor.cgi">http://pdg6.bl.gov/cgi-bin/pdg-monitor.cgi</a> for details
pdg-job-runner.sh: Wed May 30 04:05:50 PDT 2012 done.
```

```
[WARNING] /home/pdgbuild/pdgbuild-work/PdgWorkspace/src/main/java/pdg/workspace/helpermodels/EncodingTask.java:39: warning - @return tag has no arguments.
[WARNING] /home/pdgbuild-work/PdgWorkspace/src/main/java/pdg/workspace/ext/BranchingRatioFormatter.java:58: warning - @param argument "s" is not a parameter name.
[INFO] [resources:resources {execution: default-resources}]
[INFO] Using 'UTF-8' encoding to copy filtered resources.
[INFO] Copying 0 resource
[INFO] [compiler:compile {execution: default-compile}]
(INFO) Compiling 39 sounce files to /home/pdgbuild/pdgbuild-work/PdgWorkspace/target/classes
[ERROR] BUILD FAILURE
/home/pdgbuild/pdgbuild-work/PdgWorkspace/src/main/java/pdg/workspace/ext/StringFormatter.java:[128,49] cannot find symbol
symbol : method getDisplayText()
location: class pdg.macro.util.displayPrimitive.DisplayMeasurementTex
[INFO] For more information, run Maven with the -e switch
[INFO] Total time: 28 seconds
[INFO] Finished at: Wed May 30 04:05:20 PDT 2012
[INFO] Final Memory: 59M/981M
[INFO] -----
building done.
```



Cross-Linking w/PDG Identifiers



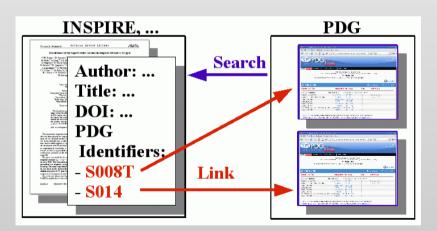
Wish list:

– From INSPIRE: "What data does PDG have about this?"

From PDG: "What are the latest papers on this topic?"

Permanent reference to PDG data items: PDG Identifiers

- Essentially PDG nodes (e.g. "Q007TP")
- Supported by pdgLive: e.g. http://pdglive.lbl.gov/view/Q007TP
- Authoritative list available from pdg.lbl.gov (under "Downloads")



Used for cross-linking to/from INSPIRE

Interest from Elsevier to do something similar in ScienceDirect



Data Entry by Collaborations?



- Should we allow the community to enter new results directly into the PDG database?
 - In particular large collaborations such as LHC experiments
 - A result is best known/understood by its author
 - PDG would still validate all entries to be published in RPP
 - Could start with a draft entry rather than just the article
 - Easier to include latest results
 - In principle, pdgLive, could show user-contributed data in addition to data vetted by PDG
 - Users choose whether they want to see user-contributed entries
 - User-contributed entries could appear online immediately
 - User-contributed data might include unpublished (public) data
 - Fits and averages including user-contributed data possible
 - Should be able to minimize abuse by requiring authentication and displaying the identity of contributiors

Is this something worth pursuing?